DOCUMENT RESUME

ED 389 713 TM 024 193

AUTHOR Wightman, Linda F.; Wightman, Lawrence E.
TITLE An Empirical Investigation of One Variable Section Pre-Equating.
INSTITUTION Educational Testing Service, Princeton, N.J.
REPORT NO ETS-RR-88-37
PUB DATE Jun 88
NOTE 44p.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Comparative Analysis; Correlation; *Equated Scores; Estimation (Mathematics); Models; *Statistical Analysis; *Test Format
IDENTIFIERS Graduate Management Admission Test; *Missing Data; *Parallel Test Forms

ABSTRACT

Section Pre-Equating (SPE) is a method used to equate test forms that consist of multiple separately timed sections. SPE does not require examinees to take two complete forms of the test. Instead, all of the old form and one or two sections of the new form are administered to each examinee, and missing data techniques are employed to estimate the necessary equating parameters. When a test includes only one variable section, there is no simple way to obtain an estimate of the correction between pairs of sections from the new test because no group takes a pair of sections. This study evaluated several methods for obtaining reasonable values for these correlations. Methods evaluated included borrowing correlations from another population that took the test at a different administration, borrowing correlation from another parallel test, and imputing the unknown partial correlations. Samples were drawn from existing data for the Graduate Management Admission Test. Comparison of the converted scores obtained using the three different methods showed them to be consistent with one another. Results also demonstrated that equating results from the one-variable section test model were very consistent with those obtained from a two-variable section test model. An appendix presents section correlations. (Contains 8 references, 10 tables, and 12 figures.) (Author/SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 389 713

RESEARCH REPORT

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

H. I. BRAUN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC).

AN EMPIRICAL INVESTIGATION OF ONE VARIABLE SECTION PRE-EQUATING

Linda F. (Leary) Wightman
Lawrence E. Wightman

Educational Testing Service
Princeton, New Jersey
June 1988

BEST COPY AVAILABLE

TM024193

An Empirical Investigation of One Variable Section Pre-Equating[1]

Linda F. (Leary) Wightman

Lawrence E. Wightman

Educational Testing Service

June 1988

---

[1]Analyses in this paper are based (in part) on data used with permission of the Graduate Management Admission Council (GMAC). The GMAC encourages researchers to formulate and freely express their own opinions, and the opinions expressed here are not necessarily those of the GMAC.

Copyright © 1988. Educational Testing Service. All rights reserved.

Abstract

Section Pre-Equating (SPE) is a method used to equate test forms that consist of multiple separately-timed sections. The unique contribution of SPE is that it does not require examinees to take two complete forms of the test. Instead, all of the old form and one or two sections of the new form are administered to each examinee. Missing data techniques are employed to estimate the necessary equating parameters. When a test has two variable sections, the estimation is fairly straightforward.

When a test includes only one variable section, there is no simple way to obtain an estimate of the correlation between pairs of sections from the new test because no one group takes a pair of sections. This study was designed to utilize empirical data to evaluate several methods for obtaining reasonable values for these correlations. The methods for obtaining correlations that are evaluated are borrowing correlations from another population who have taken the same test at a different administration, borrowing correlations from another parallel test taken by the same population, and imputing the unknown partial correlations. This study explored the effect on equating of using each of these methods. The criterion for evaluating the one-variable-section test equatings was the equating obtained using a two variable section model.

Comparison of the converted scores obtained using the three different methods for estimating correlations under a one-variable-section test model were consistent with one another. Results from this study also demonstrated that equating results obtained from a one-variable-section test model were very consistent with those obtained from a two-variable section test model.

An Empirical Investigation of One Variable Section Pre-Equating

This study was designed to utilize data gathered from several years of implementation of two-variable section pre-equating (SPE) to evaluate the potential impact on equating of replacing the two-variable-section model with a one variable section model. Development of SPE procedures for a one-variable-section test (1VST) provides the following potential advantages:

- o Shortened testing time - one less nonoperational section is administered.
- o Increased security - only half as many appearances (exposures) of pre-equating test items to an individual test taker.
- o Availability of SPE as a pre-equating option for testing programs that cannot include two nonoperational sections in their test format.

Theoretical Framework

Section Pre-Equating (SPE) is a method appropriate for equating tests which consist of multiple, separately-timed sections (Holland and Wightman, 1982). It was developed to parallel the equating concept of administering new and old forms of a test to the same population and equating the new to the old form by estimating the parameters necessary to solve the standard linear equating function. The unique contribution of SPE is that it does not require examinees to take two complete forms of the test. Instead, all sections of one form of the test are given to all examinees. The sum (or partial sums) of these sections comprise each examinee's score(s) for the test. The sections of the test to be pre-equated, referred to as the preoperational test, are introduced into the operational test through the use of variable sections. Each test version contains all sections of the

operational test plus one or two sections of the new form. The content of the variable sections differs from one test book to another. That is, different versions of the test contain the same operational sections, but different sections or combinations of sections from the preoperational test. An example of a two-variable-section test model for a test made up of three verbal sections and three quantitative sections might be:

Verbal-1 Quant-1 Verbal-2 **Variable-1** Verbal-3 Quant-2 **Variable-2** Quant-3

Thus, each examinee takes the complete operational test, but only one or two sections of the preoperational test. Missing data techniques (Holland and Thayer, 1985) are applied to estimate the parameters necessary to equate the preoperational test to the operational test. When the test has two variable sections (2VST), the estimation is fairly straightforward. Consider an operational test X with m sections (X_1, X_2, ...X_m) and a preoperational test Y, also with m sections. A random sample of examinees (P_{ij}) is tested with a version including sections Y_i and Y_j. An illustration of the sampling scheme resulting from a two-variable-section test is presented in Figure 1. Pooling all of the examinees tested with versions including Y_i yield the sample mean and variance of Y_i, which are unbiased estimates of the population mean and variance. The same is true for Y_j. To estimate the correlation rho_{ij}, data from the sample r_{ij} are used to estimate the population cross moments. Both "pairwise present" and maximum likelihood methods have been used to obtain estimates required to calculate the equating parameters.

Insert Figure 1 about here.

When the test has only one variable section (1VST), there is no simple way to obtain an estimate of the correlation between Y_i and Y_j, because no one group takes both Y_i and Y_j. An illustration of the sampling scheme resulting from a one-variable-section test is presented in Figure 2. Several methods for obtaining reasonable values for these correlations have been suggested by Holland and Wightman (1982). These include borrowing correlations from another population who have taken the same test at a different administration, borrowing correlations from another parallel test taken by the same population, and imputing the unknown partial correlations. This study explored the effect on equating of using each of these methods for obtaining unobserved test section correlations with a one-variable-section test SPE model.

Insert Figure 2 about here.

RESEARCH DESIGN

Description of the Test

This study used empirical data from the Graduate Management Admission Test (GMAT) to mimic the administration and pre-equating of a one-variable-section test. The GMAT forms used in this study were made up of three verbal sections (Analysis of Situations, Reading Comprehension, and Sentence

Correction) and three quantitative sections (two Problem Solving and Data Sufficiency). The sum of the raw scores on each verbal operational section yields a Verbal Score and, likewise, the sum of the scores on the three quantitative operational sections yields a Quantitative Score. In the standard two-variable-section test SPE administration of the GMAT, a test form includes eight sections-- three operational verbal sections, three operational quantitative sections, and two preoperational variable sections. The ordering of the preoperational verbal and quantitative sections, as well as the position of each of the variable sections differs from one administration to the next. In this study, we estimated one-variable-section test pre-equating parameters for five different forms of the GMAT. The positions of the operational and variable sections of the tests used is this study are shown in Table 1. Each of these forms had previously been pre-equated using the two variable SPE model. Forms F2 and F4 were pre-equated during the same operational administration, Form G1 was pre-equated during a different operational administration, and Forms G4 and H1 were pre-equated during a third operational administration.

Insert Table 1 about here.

Different subforms include different pairs of preoperational verbal or quantitative item sections. For a form to be used in this study, each verbal item type and each quantitative item type had to appear at least in the first variable section position.

Description of the Samples

Samples were drawn from existing GMAT data. Each sample included only test takers who took a subform that included sections from the preoperational form selected for analysis. For example, Forms F2 and F4 were administered preoperationally at one operational administration. At that same operational administration, new items were also pretested. Twenty subforms of operational Form A were administered. Six subforms included Form F2 preoperational test sections. Only the test takers who took one of those six subforms were included in the one-variable-section test equating sample used to equate F2 to A in this study. Preoperational means and variances for each of the six item types were calculated from only those test takers who took the item type in the position being considered as the 1VST variable section. This procedure was followed in order to reflect a one variable administration as closely as possible. When a 2VST is used, the item type means and variances are calculated from observed data obtained from both positions. In effect, the use of a one variable section model reduces the equating sample almost in half. Table 2 shows the sample sizes used to perform each of the one variable equatings.

Insert Table 2 about here.

Methods

This study capitalized on the availability of several years of empirical data to provide baseline information about the correlations among the

different verbal and quantitative items types. The linear equating parameters obtained from the two variable section pre-equating were used as the criterion against which each of three one variable section pre-equatings were evaluated.

In order to estimate a one variable section test pre-equating situation, one of the two variable sections from several GMAT test forms was ignored and the tests were re-equated. For example, in one of the test forms used in this study, the variable sections appeared in positions 3 and 8. All response data from section 8 were ignored and new equating parameters were estimated based only on data from sections 1 through 7. A second pre-equating was then estimated by ignoring the data in section 3 and using only the data from sections 1, 2, 4, 5, 6, 7, and 8. Differences between first and second variable position equatings would reflect the impact of section position on equating. This was of particular concern in this study because a practice effect resulting from the two-variable-section test pre-equating design had been established in earlier SPE studies (Wightman, 1981; Wightman and Leary, 1983; Faggen and McPeek, 1981).

One criterion used in selecting forms for analysis in this study was that each of the six operational item types must appear in the variable section position under consideration. In the early administrations of the GMAT using two-variable-section test SPE, this constraint was always met. In later administrations, all verbal and quantitative sections do not appear in both variable positions. If each of the six test sections did not appear in a variable position in a test form, that form (or position within that form) could not be used to mimic a one-variable-section test administration. This limited the number of test forms that could be used in this study.

For Forms F2, F4, G4, and H1, each item type appeared in the variable section in the first position and also appeared in the variable section in the second position. For these data, the one variable SPE parameters were estimated twice (first assuming that the first position was the one variable section, and ignoring the data in the second variable position, and then assuming that the second position was the one variable section and ignoring the data in the first variable position). Analysis of data from Form G1 was limited to consideration of the first variable section. Only five of the six item types appeared in the second variable position.

The unobserved covariances for the one-variable-section test pre-equatings in this study were estimated using each of three methods. Each method is described separately.

Borrowing Correlations from Another Test. The operational test X and the preoperational test Y are designed to be parallel in content and difficulty. Assuming that they are, in fact, parallel, we borrowed the observed correlations between sections from the operational test X to estimate the corresponding unobserved covariances for the preoperational test Y. That is,

$$cov_{ij} = r_{i'j'} s_i s_j \quad (1)$$

where cov_{ij} = estimated covariance between section i and section j in the preoperational test

$r_{i'j'}$ = observed correlation between section i and section j in the operational test

s_i, s_j = observed standard deviations for sections i and j in the preoperational test

Borrowing Correlations from Another Population. Each of the test forms used in this study was administered preoperationally on one date to one population of test takers and was administered operationally at a later date to a different population. Assuming that

$$\text{rho } Y_i Y_j(\text{preoperational population}) = \text{rho } Y_i Y_j(\text{operational population}), \quad (2)$$

we equated **Y** to **X** by borrowing the observed correlations from the later operational administration of **Y**.

Imputing the Unknown Partial Correlations. In the one variable section pre-equating situation, it is the partial correlation between Y_i and Y_j, after partialling out X_1, X_2, ...X_m, that is inestimable. Rubin and Thayer (1978) studied the sensitivity of estimates of the variance/covariance matrix by varying the unknown partial correlations over a plausible range and combining these values with estimates of the estimable parameters in the matrix. We used the Rubin and Thayer methodology in this study. For each verbal equating estimated in this study, the six operational sections were taken by every test taker, but each preoperational section was taken by non-overlapping samples of test takers. The method proposed by Rubin and Thayer to calculate the sample covariance matrix necessary to equate the preoperational test to the operational test is to isolate those parameters about which the data supply no information. The parameters about which there is no information are the partial correlations among the preoperational sections (**Y** variables) given the parallel operational sections (**X** variables). Once these parameters are isolated, plausible values can be inserted for them and maximum likelihood estimates (MLE) can be inserted for the other values. Note that unique maximum likelihood estimates exist for the mean vector and covariance matrix of the operational sections (X_1, X_2, and X_3), for the regression of the

preoperational sections on the operational sections (Y on X), and for the residual variances of Y given X . By inserting plausible values for the partial correlations among the preoperational sections (Y) and unconditioning on X , we obtained maximum likelihood estimates for the mean vector and covariance matrix of (X,Y).

For this study, we used the minimum, maximum, and median observed values of the partial correlations between $Y_i Y_j$ preoperational sections, controlling for the operational sections, obtained over multiple forms of this test, as plausible values for the inestimable parameters.

Estimating Equating Parameters. Standard linear equating parameters were estimated by setting the MLE mean and standard deviation of each preoperational form equal to the observed unpracticed mean and standard deviation of the corresponding operational form. The equating obtained using the two-variable-section test pre-equating methodology was used as the criterion against which to evaluate each of the one-variable-section test covariance estimation procedures.

RESULTS

The preoperational verbal and quantitative scores from each of five GMAT forms were equated to the operational verbal and quantitative scores three times using each of the three methods described earlier to estimate linear equating parameters under a one-variable-section test model. The scores from each form were then placed on the GMAT scale using the linear scaling parameters extant for each corresponding operational test. All equatings for

Forms F2, F4, G4, and H1 were done twice, once using preoperational data from the first variable section position, then using data from the second variable section position.

There is no completely satisfactory criterion for evaluating the correctness of a one-variable-section test equating. For purposes of this study, we used the two-variable-section test equating as a criterion. If the one-variable-section test results in an equating that is virtually indistinguishable from the two-variable-section test equating, there is some confirmation that each method is producing an essentially correct result. The appropriateness of using the two-variable-section test SPE as a criterion is strengthened by the results of a study comparing the two variable SPE method with an IRT equating method (Kingston, Leary, and Wightman, 1985). The results of the IRT/SPE comparison suggested that the equatings from the two methods were consistent for both the Verbal and the Quantitative measures.

Selecting Partial Correlations

As noted earlier, the major difficulty in using a one-variable-section test model is that the partial correlation between Yi and Yj is inestimable. Empirical data gathered over the past several years were used to determine a reasonable range for these partial correlations. This was accomplished by calculating the observed partial correlation between each pair of verbal sections after partialling out all of the operational verbal sections; the process was repeated for each quantitative pair. Partial correlations were estimated in this way for each of the three verbal pairings and each of the three quantitative pairings for each of ten independent observations of GMAT data. From these data, the minimum, maximum, and median partial correlation

values were selected for each preoperational section pairing. These values are shown in Table 3. The observed partial correlations showed little variation across the several test forms. The largest observed difference between the minimum and maximum partial correlations was .20 .

Insert Table 3 about here.

Comparison of Equatings Obtained from Using Different Partials

A separate verbal and quantitative variance/covariance matrix using each set of partial correlations shown in Table 3 was estimated for each of the preoperational forms. As a further estimate of the sensitivity of varying the unknown partials, a fourth variance/covariance matrix was estimated for each form using a fixed partial correlation value of .10 for each section pair. One variable SPE linear equating parameters were calculated for each test form (Verbal and Quantitative) by setting the preoperational mean and standard deviation equal to the operational mean and standard deviation. The preoperational scores were placed on the GMAT scales through the linear scaling parameters for the corresponding operational form. Tables 4 and 5 present unrounded equated scaled scores obtained using each of the four sets of imputed partial correlations for selected raw scores for the Verbal and Quantitative scores, respectively. The tables also include the results of the two variable SPE for each test form.

The data in Tables 4 and 5 do not show any set of partials to be clearly superior or inferior. That is, the equatings do not seem to be particularly

sensitive to the use of different reasonable values of the unknown partial correlations. The median partial correlations tended to result in conversion lines that are closest to the two variable SPE estimates. The MLE variance/covariance matrices resulting from the use of the median partial correlations were used to compare the equating parameters for the imputed partial correlations 1VST method with the equating results obtained using other 1VST methods discussed in the remainder of this paper.

Comparison of Equating Results Among the 1VST Methods

Tables 6, 7, 8, and 9 show 1VST unrounded converted scores for each of the different 1VST methods (imputed partial correlations, borrowed from a different test, borrowed from a different population) and 2VST unrounded converted scores for each of the five GMAT forms. Table 6 shows verbal converted scores using data from the first variable section position; Table 7 shows verbal converted scores using data from the second variable section position. Likewise, Tables 8 and 9 show Quantitative converted scores using data from the first and second variable section positions, respectively.

Insert Tables 6 through 9 about here.

One striking feature of these data is the differences in the resulting 1VST converted scores between the first and second variable section positions. Converted scores estimated from preoperational test sections appearing in the second variable position were consistently lower than the same scores estimated from preoperational test sections appearing in the first variable position. These results are consistent with the results of several practice

effects studies cited earlier. That is, test takers seem to benefit from the opportunity to practice and a test section that has been preceded by a section of the same item type appears easier than the same section when it appears in an earlier test section where there is less opportunity for practice. The 2VST SPE method takes the effect of practice into account by separately estimating means and variances for practiced and for unpracticed data. The equating parameters for the unpracticed two-variable-section test are used in operational GMAT SPE and are, therefore, the parameters that were used as the criterion in this study. The preoperational data from each 1VST method were equated to operational data that ignored the effect of practice. The larger difference between the one-variable-section equatings using data from the second variable section position confirms the presence of practice effects, their effect on equating, and the appropriateness of correcting for practice in the 2VST SPE model.

The differences between the 1VST and the 2VST equatings are consistently small. Table 10 shows the maximum converted score differences between the estimated 1VST conversion function and the criterion 2VST conversion function for each of the three 1VST methods. Comparison of these differences suggests that, given the two-variable-section test linear equating as the criterion, a consistent one-variable-section test equating results from application of any of the 1VST methods. The largest differences between the 2VST equating and each of the 1VST equatings is seen in the verbal equating of Form G1. The variable section in this form appeared in Section 1. Thus all of the preoperational data are unpracticed. One consequence of preoperational sections appearing early in the test is increased opportunity for practice on the operational sections. Empirical evidence suggests that performance on

verbal sections benefits more from practice than performance on quantitative sections. This effect of practice is evident in the verbal equating results for Form G1.

Insert Table 10 about here.

In general, the equatings resulting from the use of imputed partial correlations produce conversion lines slightly closer to the 2VST conversion lines than equatings resulting from borrowing correlations from a different test. This result is consistent with results reported in an earlier study by Holland and Wightman (1982), in which the SPE data were simulated from SAT data. However, the differences among the 1VST methods found in the GMAT data are much smaller than those found by Holland and Wightman. The smaller differences may be attributed at least partially to the fact that the GMAT data were developed for and administered under an SPE model. Unlike the results reported by Holland and Wightman, equatings using correlations borrowed from a different population were as consistent as the equatings using imputed partial correlations. The disadvantage to the operational implementation of borrowing from a different population is that it depends on the operational administration in order to complete the equating. Using imputed partial correlations allows continuation of the current practice of pre-equating.

A more pragmatic way of viewing the impact of the 1VST methods is in terms of score reporting practices. GMAT uses formula scoring, that is, number right minus 1/4 number wrong rounded to the nearest integer. For the verbal measure, this introduces differences in scaled scores between rounded

and unrounded raw formula scores of as much as .3 scaled score points. In addition, scaled scores are rounded to the nearest integer for score reporting. Thus, for any given raw score, the reported score might be one scaled score point different than the unrounded equated score. Table 10 shows that the maximum difference between the 1VST and the 2VST equatings never exceeds one rounded scaled score point.

Figures 3 through 7 show graphically the results of the Verbal equatings using sections in the first variable position. Figure 3 shows four conversion lines for Form F2--(1) the 2VST conversion line, (2) the 1VST conversion line from borrowing correlations from a different test, (3) the 1VST conversion line from borrowing correlations from a different population, and (4) the 1VST conversion line from imputing partial correlations. Figures 4 through 7 show these same conversion lines for Forms F4 through H1, respectively. Figures 8 through 12 show the results of the Quantitative equatings for the same five forms, in the same way. In each figure, the conversion lines are so close to each other that they are, for the most part, indistinguishable.

Insert Figures 3 through 12 about here.

Summary of Equating Results

Comparison of the converted scores for the three different 1VST methods shows that the equatings were very consistent for both the Verbal and the Quantitative measures of the GMAT. The converted scores from any of the 1VST equating methods did not vary from the converted scores obtained using a 2VST section pre-equating method by more than one scaled score point at any point

along the raw score scale. Because the differences in the GMAT reported scores derived using standard rounding practices result in differences of this magnitude, the differences between 1VST and 2VST equating methods are judged to be negligible.

One concern about the interpretation of these data and the recommendation to adopt the use of operational SPE is related to the effect of practice on the equatings. The equatings performed for this study confirmed that practice does indeed have an effect. These results are consistent with the findings of Holland and Thayer (1981). When they used SPE to equate the GRE General Test, the practice effect known to affect the GRE analytical sections resulted in a marked bias in the SPE results.

The data used in this study mimicked a 1VST administration and equating by ignoring one of the two variable sections in the two variable section test. Although the data from one variable section were ignored, any benefits of practice from the ignored section that were accrued in the succeeding operational sections could not be eliminated. In a two-variable-section test SPE, the program procedure is to account for practice and equate only on unpracticed data. The effects of practice on the GMAT equatings might be diminished in a real one-variable-section test administration. That is, in a given test booklet only one item type would be potentially affected by practice. The actual benefits cannot be evaluated unless a true one-variable-section test model is administered. The results of this study support consideration of the adoption of a one-variable-section test equating model.

REFERENCES

Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977) Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society (B), 39, 1-38.

Faggen, J. & McPeek, M. (1981) Practice effects for four different item types. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles.

Holland, P. W., & Thayer, D. T. (1981) Section Pre-Equating the Graduate Record Examination (Program Statistics Research Report No. 81-13). Princeton, NJ: Educational Testing Service.

Holland, P. W., & Thayer, D. T. (1985) Section Pre-Equating in the presence of practice effects. Journal of Educational Statistics, 10, 109-120.

Holland, P. W. & Wightman, L. E. (1982) Section Pre-Equating: A preliminary investigation. In P. W. Holland & D. B. Rubin (Eds.), Test Equating (pp. 271-297). New York: Academic Press.

Rubin, D. B., & Thayer, D. T. (1978) Relating tests given to different samples. Psychometrika, 43, 3-10.

Wightman, L. E. (1981) GMAT within-test practice effects studies. Paper presented at the annual meeting of the National Council on Measurement in Education, Los Angeles.

Wightman, L. E. & Leary, L. F. (1983) Dealing with practice effects in section pre-equating. Paper presented at the annual meeting of the American Educational Research Association, Montreal.

Table 1

Section Positions for Five Preoperational Forms of the GMAT

Form	Administration Date	Section	Item Type	Section	Item Type
F2 & F4	6/82	1	Reading Comprehension	5	Data Sufficiency
		2	Variable	6	Sentence Correction
		3	Problem Solving 1	7	Variable
		4	Analysis of Situations	8	Problem Solving 2
G1	10/82	1	Variable	5	Variable
		2	Problem Solving 1	6	Analysis of Situations
		3	Sentence Correction	7	Problem Solving 2
		4	Data Sufficiency	8	Reading Comprehension
G4 & H2	1/84	1	Reading Comprehension	5	Data Sufficiency
		2	Problem Solving 1	6	Analysis of Situations
		3	Variable	7	Problem Solving 2
		4	Sentence Correction	8	Variable

Table 2

Numbers of Candidates for Test Subforms

Showing Pairs of Preoperational Sections

Preop Form	First Position Section	Item Type	Second Position Section	Item Type	N
F2	2	Reading Comp.	7	Anal.of Sit.	2,396
	2	Anal. of Sit.	7	Sent. Corr.	2,412
	2	Sent. Corr.	7	Reading Comp.	2,483
	2	Prob. Solv. 1	7	Data Suff.	2,361
	2	Prob. Solv. 2	7	Prob. Solv. 1	2,446
	2	Data Suff.	7	Prob. Solv. 2	2,397
F4	2	Reading Comp.	7	Anal. of Sit.	2,349
	2	Anal. of Sit.	7	Sent. Corr.	2,314
	2	Sent. Corr.	7	Reading Comp.	2,454
	2	Prob. Solv. 1	7	Data Suff.	2,483
	2	Prob. Solv. 2	7	Prob. Solv. 1	2,387
	2	Data Suff.	7	Prob. Solv. 2	2,349
G1	1	Reading Comp.	5	Not all item types observed; therefore, not used in second position.	2,832
	1	Anal. of Sit.	5		2,806
	1	Sent. Corr.	5		2,862
	1	Prob. Solv. 1	5		2,951
	1	Prob. Solv. 2	5		2,810
	1	Data Suff. position.	5		2,774
G4	3	Reading Comp.	8	Anal. of Sit.	2,112
	3	Anal. of Sit.	8	Sent. Corr.	2,285
	3	Sent. Corr.	8	Reading Comp.	2,204
	3	Prob. Solv. 1	8	Prob. Solv. 2	2,238
	3	Prob. Solv. 2	8	Data Suff.	2,207
	3	Data Suff.	8	Prob. Solv. 1	2,143
H2	3	Reading Comp.	8	Anal. of Sit.	2,176
	3	Anal. of Sit.	8	Sent. Corr.	2,145
	3	Sent. Corr.	8	Reading Comp.	2,197
	3	Prob. Solv. 1	8	Prob. Solv. 2	2,295
	3	Prob. Solv. 2	8	Data Suff.	2,207
	3	Data Suff.	8	Prob. Solv. 1	2,159

Table 3

Partial Correlations Used in the Study

Verbal Preoperational Item Type Pairs

and Quantitative Preoperational Item Type Pairs

Verbal Pairs			
	AS/RC	AS/SC	RC/SC
Minimum:	.101	.098	.120
Median:	.145	.124	.179
Maximum:	.237	.175	.219
Quantitative Pairs			
	PS1/PS2	PS1/DS	PS2/DS
Minimum:	.144	.129	.134
Median:	.284	.200	.184
Maximum:	.339	.258	.286

Table 4

Converted Verbal Scores

from One Variable Section Equatings

using Partial Corelations

		Selected Raw Scores								
Form	Partials	0	10	20	30	40	50	60	70	80
F2	rho=.10	4.82	10.36	15.89	21.42	26.96	32.49	38.02	43.56	49.09
	rho=min	4.84	10.37	15.90	21.43	26.96	32.48	38.01	43.54	49.07
	rho=median	5.00	10.49	15.97	21.46	26.95	32.43	37.92	43.41	48.89
	rho=max	5.23	10.66	16.08	21.51	26.94	32.37	37.80	43.22	48.65
	2VST SPE	4.78	10.33	15.87	21.41	26.96	32.50	38.05	43.59	49.13
F4	rho=.10	2.99	9.24	15.49	21.74	27.98	34.23	40.48	46.73	52.98
	rho=min	3.01	9.25	15.50	21.74	27.98	34.23	40.47	46.71	52.96
	rho=median	3.19	9.40	15.60	21.81	28.01	34.22	40.42	46.63	52.83
	rho=max	3.44	9.56	15.69	21.82	27.94	34.07	40.19	46.32	52.45
	2VST SPE	3.61	9.72	15.83	21.94	28.05	34.16	40.27	46.38	52.49
G1	rho=.10	4.40	10.69	16.97	23.25	29.54	35.82	42.11	48.39	54.67
	rho=min	4.43	10.70	16.98	23.26	29.53	35.81	42.09	48.37	54.64
	rho=median	4.62	10.84	17.07	23.29	29.52	35.75	41.97	48.20	54.42
	rho=max	4.89	11.04	17.19	23.35	29.50	35.65	41.80	47.95	54.10
	2VST SPE	4.01	10.19	16.37	22.55	28.73	34.91	41.09	47.27	53.45
G4	rho=.10	1.92	7.97	14.01	20.06	26.11	32.16	38.21	44.26	50.31
	rho=min	1.94	7.98	14.03	20.07	26.11	32.16	38.20	44.24	50.29
	rho=median	2.13	8.13	14.13	20.12	26.12	32.12	38.12	44.11	50.11
	rho=max	2.41	8.34	14.27	20.20	26.13	32.06	37.99	43.92	49.85
	2VST SPE	2.14	8.10	14.07	20.03	25.99	31.96	37.92	43.88	49.85
H1	rho=.10	5.61	11.30	16.98	22.67	28.36	34.05	39.74	45.42	51.11
	rho=min	5.63	11.31	17.00	22.68	28.36	34.04	39.72	45.41	51.09
	rho=median	5.81	11.45	17.08	22.72	28.35	33.99	39.63	45.26	50.90
	rho=max	6.06	11.63	17.20	22.77	28.34	33.91	39.48	45.05	50.62
	2VST SPE	5.68	11.28	16.89	22.49	28.10	33.70	39.31	44.91	50.52

Table 5

Converted Quantitative Scores

from One Variable Section Equatings

using Partial Corelations

Form	Partials	Selected Raw Score 0	10	20	30	40	50	60
F2	rho=.10	9.62	16.44	23.27	30.09	36.91	43.73	50.55
	rho-min	9.71	16.50	23.29	30.07	36.86	43.64	50.43
	rho-median	9.93	16.63	23.33	30.04	36.74	43.44	50.14
	rho-max	10.11	16.74	23.37	30.01	36.64	43.28	49.91
	2VST SPE	9.74	16.46	23.19	29.92	36.65	43.37	50.10
F4	rho=.10	9.71	16.18	22.66	29.13	35.60	42.08	48.55
	rho-min	9.80	16.24	22.68	29.13	35.57	42.01	48.45
	rho-median	10.00	16.37	22.74	29.10	35.47	41.84	48.20
	rho-max	10.17	16.48	22.78	29.09	35.39	41.70	48.01
	2VST SPE	9.84	16.15	22.46	28.77	35.08	41.39	47.70
G1	rho=.10	8.99	15.48	21.97	28.46	34.95	41.44	47.93
	rho-min	9.08	15.54	22.00	28.46	34.92	41.38	47.84
	rho-median	9.29	15.68	22.06	28.45	34.84	41.22	47.61
	rho-max	9.47	15.79	22.12	28.44	34.77	41.09	47.42
	2VST SPE	9.01	15.40	21.79	28.17	34.56	40.95	47.33
G4	rho=.10	9.68	16.70	23.72	30.73	37.75	44.77	51.79
	rho-min	9.57	16.55	23.53	30.51	37.49	44.48	51.46
	rho-median	9.31	16.20	23.09	29.99	36.88	43.78	50.67
	rho-max	9.09	15.91	22.73	29.55	36.37	43.20	50.02
	2VST SPE	9.24	16.11	22.99	29.86	36.73	43.61	50.48
H1	rho=.10	7.67	14.48	21.30	28.12	34.93	41.75	48.56
	rho-min	7.78	14.56	21.34	28.12	34.91	41.69	48.47
	rho-median	8.03	14.74	21.44	28.14	34.84	41.54	48.25
	rho-max	8.25	14.88	21.52	28.15	34.79	41.42	48.06
	2VST SPE	8.22	14.89	21.57	28.25	34.92	41.60	48.27

Table 6

Verbal Converted Scores for Three One-SPE Equatings

with Sections in the First Position

Form	Equating	Selected Raw Scores								
		0	10	20	30	40	50	60	70	80
F2	Median partials	5.00	10.49	15.97	21.46	26.95	32.43	37.92	43.41	48.89
	Borr. test	4.84	10.37	15.90	21.43	26.95	32.48	38.01	43.53	49.06
	Borr. populatn	4.84	10.38	15.92	21.45	26.99	32.52	38.06	43.60	49.13
	2VST SPE	4.78	10.33	15.87	21.41	26.96	32.50	38.05	43.59	49.13
F4	Median partials	3.19	9.40	15.60	21.81	28.01	34.22	40.42	46.63	52.83
	Borr. test	2.93	9.20	15.46	21.73	27.99	34.26	40.52	46.79	53.05
	Borr. populatn	2.93	9.19	15.44	21.70	27.95	34.21	40.46	46.72	52.97
	2VST SPE	3.61	9.72	15.83	21.94	28.05	34.16	40.27	46.38	52.49
G1	Median partials	4.62	10.84	17.07	23.29	29.52	35.75	41.97	48.20	54.42
	Borr. test	4.18	10.52	16.87	23.21	29.56	35.90	42.24	48.59	54.93
	Borr. populatn	4.18	10.44	16.69	22.95	29.21	35.46	41.72	47.98	54.24
	2VST SPE	4.01	10.19	16.37	22.55	28.73	34.91	41.09	47.27	53.45
G4	Median partials	2.13	8.13	14.13	20.12	26.12	32.12	38.12	44.11	50.11
	Borr. test	1.79	7.87	13.95	20.03	26.11	32.18	38.26	44.34	50.42
	Borr. populatn	1.79	7.83	13.87	19.92	25.96	32.00	38.04	44.08	50.13
	2VST SPE	2.14	8.10	14.07	20.03	25.99	31.96	37.92	43.88	49.85
H1	Median partials	5.81	11.45	17.08	22.72	28.35	33.99	39.63	45.26	50.90
	Borr. test	5.81	11.44	17.08	22.72	28.36	33.99	39.63	45.27	50.90
	Borr. populatn	5.81	11.44	17.08	22.71	28.35	33.98	39.62	45.25	50.89
	2VST SPE	5.68	11.28	16.89	22.49	28.10	33.70	39.31	44.91	50.52

Table 7

Verbal Converted Scores for Three One-SPE Equatings

with Sections in the Second Position

Form	Equating	Selected Raw Scores 0	10	20	30	40	50	60	70	80
F2	Median partials	4.12	9.56	15.01	20.46	25.91	31.35	36.80	42.25	47.69
	Borr. test	3.95	9.44	14.93	20.42	25.91	31.39	36.88	42.37	47.86
	Borr. populatn	3.95	9.45	14.95	20.44	25.94	31.43	36.93	42.43	47.92
	2VST SPE	4.78	10.33	15.87	21.41	26.96	32.50	38.05	43.59	49.13
F4	Median partials	3.00	9.15	15.29	21.44	27.59	33.73	39.88	46.03	52.18
	Borr. test	2.79	8.99	15.20	21.40	27.60	33.81	40.01	46.22	52.42
	Borr. populatn	2.79	8.98	15.18	21.37	27.56	33.76	39.95	46.15	52.34
	2VST SPE	3.61	9.72	15.83	21.94	28.05	34.16	40.27	46.38	52.49
G1	Form G1 did not have a complete set of preequating sections in the second variable position.									
G4	Median partials	2.01	7.88	13.76	19.64	25.52	31.40	37.27	43.15	49.03
	Borr. test	1.72	7.66	13.61	19.55	25.50	31.45	37.39	43.34	49.28
	Borr. populatn	1.72	7.63	13.53	19.44	25.35	31.26	37.17	43.08	48.99
	2VST SPE	2.14	8.10	14.07	20.03	25.99	31.96	37.92	43.88	49.85
H1	Median partials	4.91	10.46	16.01	21.56	27.12	32.67	38.22	43.77	49.32
	Borr. test	4.64	10.26	15.88	21.50	27.11	32.73	38.35	43.97	49.59
	Borr. populatn	4.64	10.26	15.88	21.49	27.11	32.73	38.34	43.96	49.58
	2VST SPE	5.68	11.28	16.89	22.49	28.10	33.70	39.31	44.91	50.52

Table 8

Quantitative Converted Scores for Three One-SPE Equatings with Sections in the First Position

Form	Equating	Selected Raw Scores 0	10	20	30	40	50	60
F2	Median partials	9.93	16.63	23.33	30.04	36.74	43.44	50.14
	Borr. test	9.71	16.50	23.29	30.07	36.86	43.65	50.44
	Borr. populatn	10.01	16.68	23.35	30.03	36.70	43.37	50.04
	2VST SPE	9.74	16.46	23.19	29.92	36.65	43.37	50.10
F4	Median partials	10.00	16.37	22.74	29.10	35.47	41.84	48.20
	Borr. test	9.72	16.19	22.66	29.14	35.61	42.08	48.55
	Borr. populatn	10.16	16.47	22.78	29.09	35.40	41.70	48.01
	2VST SPE	9.84	16.15	22.46	28.77	35.08	41.39	47.70
G1	Median partials	9.29	15.68	22.06	28.45	34.84	41.22	47.61
	Borr. test	8.90	15.42	21.94	28.46	34.99	41.51	48.03
	Borr. populatn	9.38	15.73	22.09	28.45	34.80	41.16	47.52
	2VST SPE	9.01	15.40	21.79	28.17	34.56	40.95	47.33
G4	Median partials	9.31	16.20	23.09	29.99	36.88	43.78	50.67
	Borr. test	9.23	16.04	22.86	29.67	36.48	43.30	50.11
	Borr. populatn	9.07	15.94	22.80	29.67	36.54	43.41	50.28
	2VST SPE	9.24	16.11	22.99	29.86	36.73	43.61	50.48
H1	Median partials	8.03	14.74	21.44	28.14	34.84	41.54	48.25
	Borr. test	8.25	14.88	21.52	28.15	34.79	41.42	48.05
	Borr. populatn	8.08	14.77	21.46	28.14	34.83	41.52	48.20
	2VST SPE	8.22	14.89	21.57	28.25	34.92	41.60	48.27

Table 9

Quantitative Converted Scores for Three One-SPE Equatings

with Sections in the Second Position

Form	Equating	Selected Raw Scores 0	10	20	30	40	50	60
F2	Median partials	9.55	16.19	22.82	29.45	36.09	42.72	49.36
	Borr. test	9.24	15.99	22.74	29.49	36.24	42.99	49.74
	Borr. populatn	9.54	16.18	22.82	29.46	36.10	42.73	49.37
	2VST SPE	9.74	16.46	23.19	29.92	36.65	43.37	50.10
F4	Median partials	9.54	15.88	22.23	28.57	34.91	41.26	47.60
	Borr. test	9.24	15.69	22.14	28.59	35.05	41.50	47.95
	Borr. populatn	9.70	15.99	22.27	28.56	34.84	41.13	47.42
	2VST SPE	9.84	16.15	22.46	28.77	35.08	41.39	47.70
G1	Form G1 did not have a complete set of preequating sections in the second variable position.							
G4	Median partials	9.12	15.88	22.63	29.39	36.15	42.90	49.66
	Borr. test	9.42	16.10	22.77	29.44	36.12	42.79	49.47
	Borr. populatn	9.26	15.99	22.72	29.45	36.17	42.90	49.63
	2VST SPE	9.24	16.11	22.99	29.86	36.73	43.61	50.48
H1	Median partials	7.88	14.55	21.23	27.91	34.58	41.26	47.93
	Borr. test	8.13	14.72	21.32	27.92	34.52	41.12	47.72
	Borr. populatn	7.96	14.61	21.26	27.91	34.56	41.22	47.87
	2VST SPE	8.22	14.89	21.57	28.25	34.92	41.60	48.27

Table 10

Maximum Scaled Score Differences between 1VST Methods of Conversion Function Estimation and 2VST SPE Results

Form	Imputed Partial	Borrowed Correlations	
		Test	Population
Verbal Score (1st position)			
F2	-.27	-.08	-.04
F4	-.42	-.67	-.64
G1	1.0	1.57	1.15
G4	.28	.63	.48
H1	.39	.40	.39
Quantitative Score (1st position)			
F2	.19	.37	.27
F4	.53	.93	.32
G1	.28	.76	.36
G4	.20	-.40	-.21
H2	-.18	-.24	-.14

Figure 1

Representation of a Two-variable

SPE Data Collection Design

	Operational						Preoperational					
Sample	AS	RC	SC	PS1	PS2	DS	AS	RC	SC	PS1	PS2	DS
1	X	X	X	X	X	X	X	X				
2	X	X	X	X	X	X	X		X			
3	X	X	X	X	X	X		X	X			
4	X	X	X	X	X	X				X	X	
5	X	X	X	X	X	X					X	X
6	X	X	X	X	X	X				X		X

Figure 2

Representation of a One-variable

SPE Data Collection Design

	Operational						Preoperational					
Sample	AS	RC	SC	PS1	PS2	DS	AS	RC	SC	PS1	PS2	DS
1	X	X	X	X	X	X	X					
2	X	X	X	X	X	X		X				
3	X	X	X	X	X	X			X			
4	X	X	X	X	X	X				X		
5	X	X	X	X	X	X					X	
6	X	X	X	X	X	X						X

Figure 3

GMAT Verbal
Conversion Lines From Four Equating Methods
TEST=Form 3FBS2

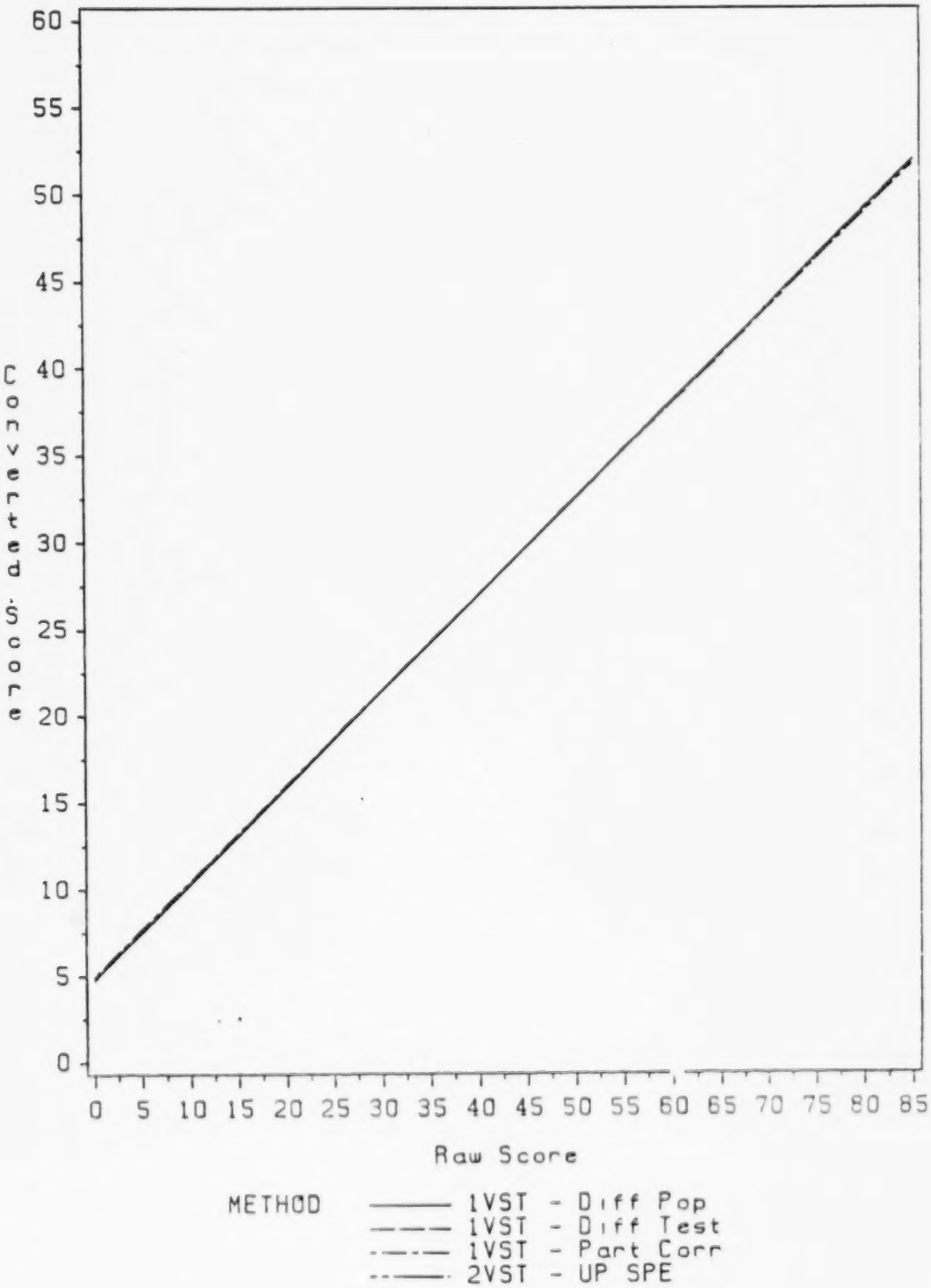


Figure 4

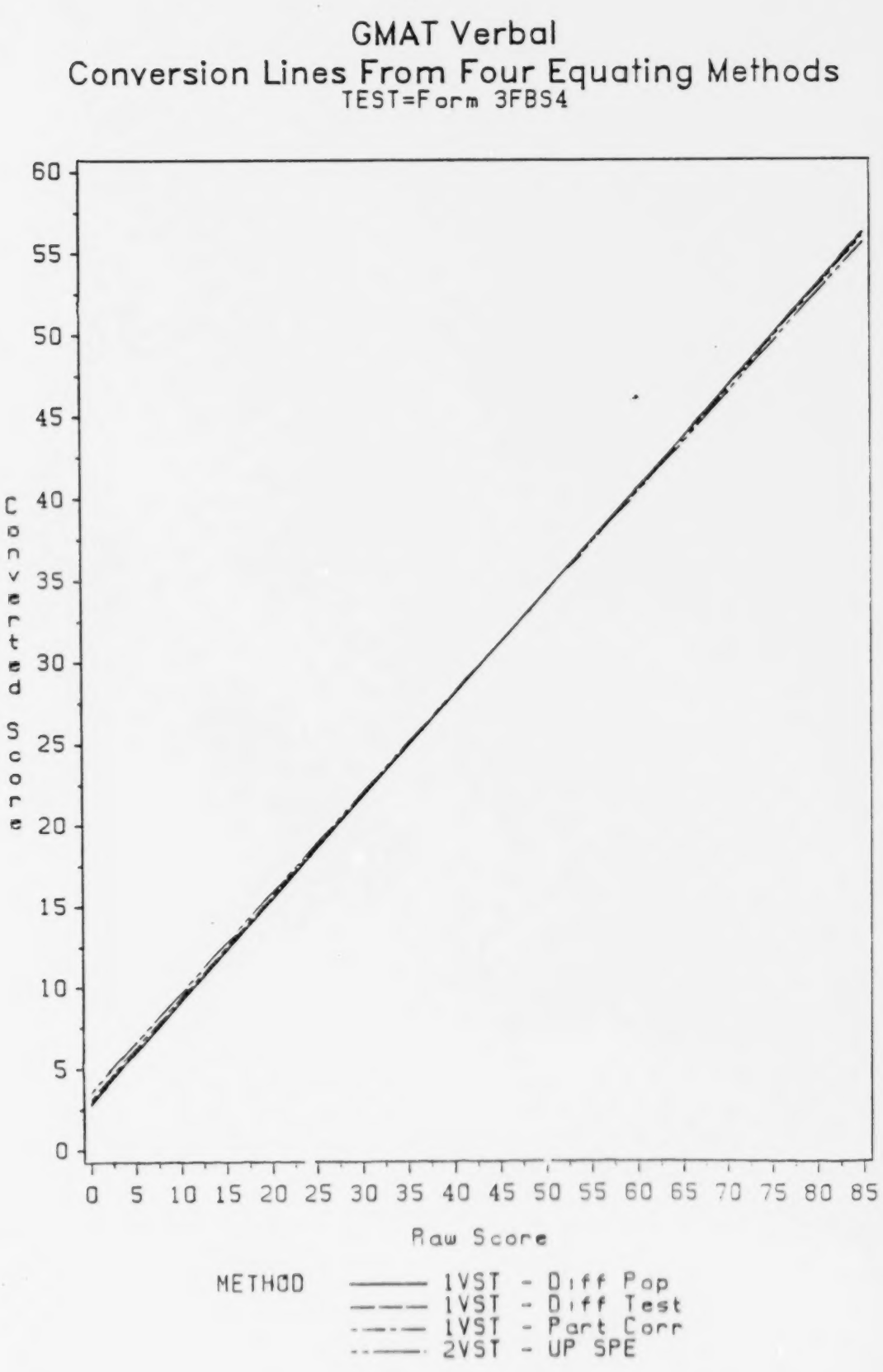

GMAT Verbal
Conversion Lines From Four Equating Methods
TEST=Form 3FBS4
Converted Score
60
55
50
45
40
35
30
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85
Raw Score
METHOD
1VST - Diff Pop
1VST - Diff Test
1VST - Part Corr
2VST - UP SPE

Figure 5

GMAT Verbal
Conversion Lines From Four Equating Methods
TEST=Form 3GBS1

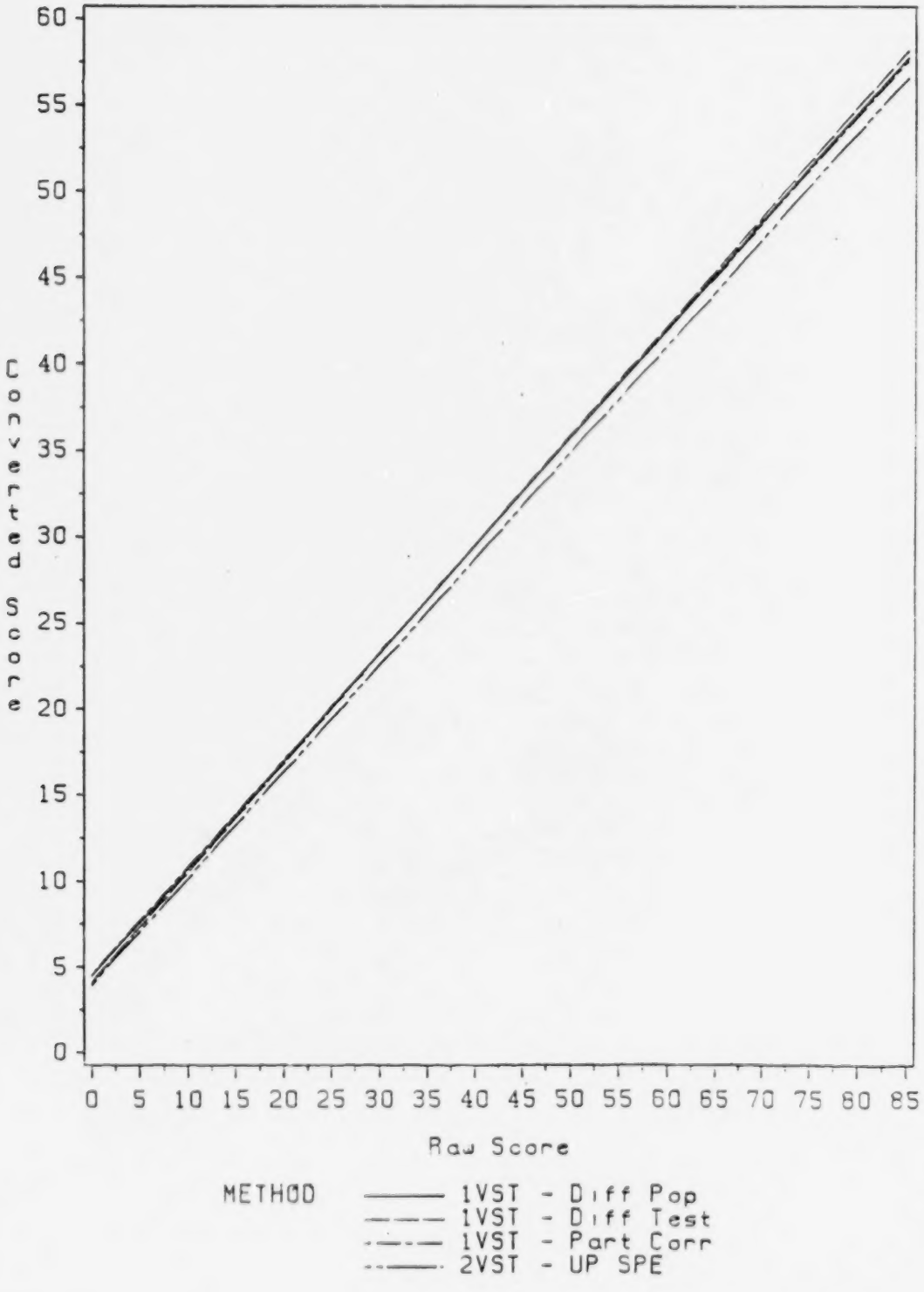


Figure 6

GMAT Verbal
Conversion Lines From Four Equating Methods
TEST=Form 3GBS4

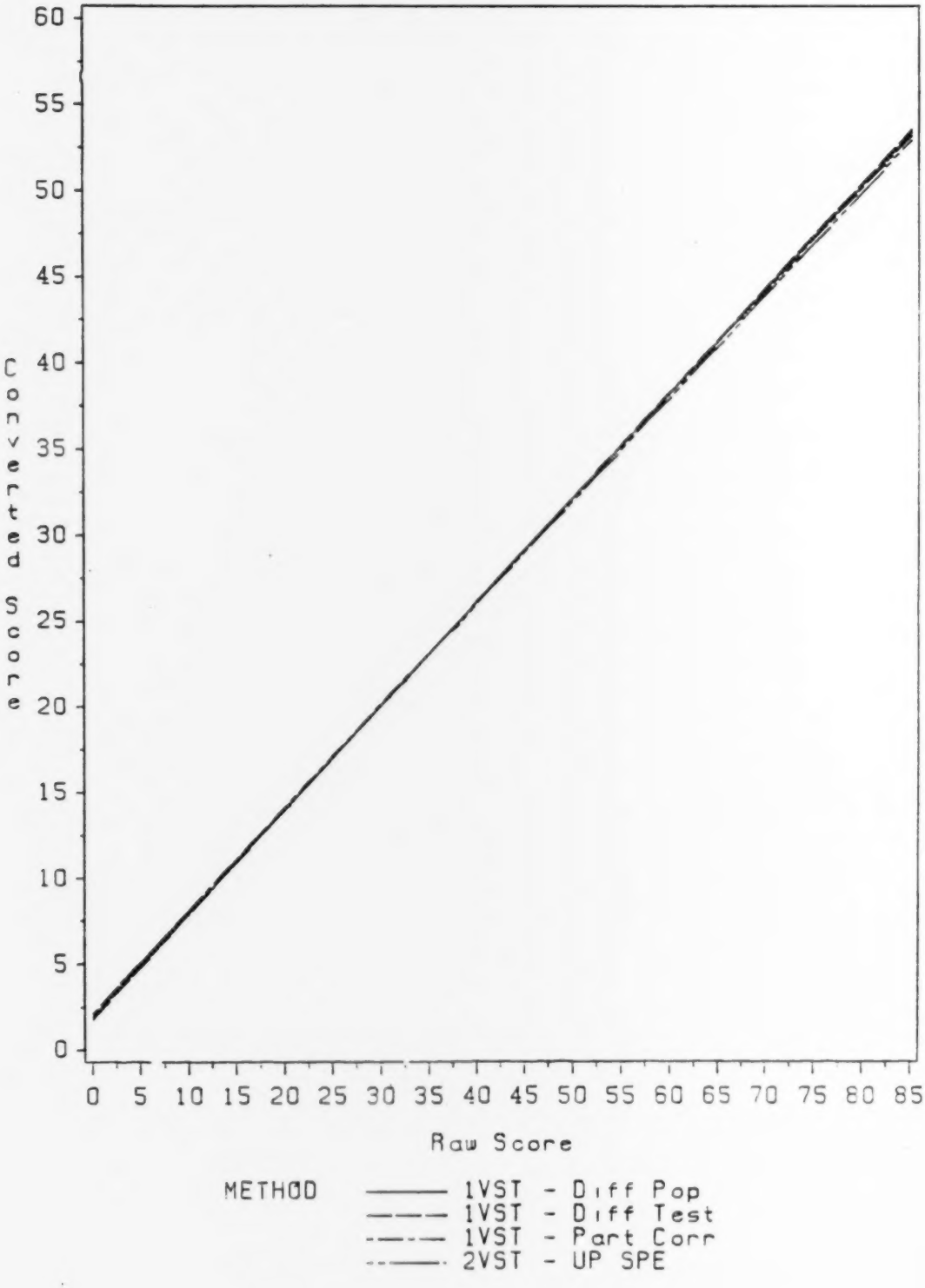


Figure 7

GMAT Verbal
Conversion Lines From Four Equating Methods
TEST=Form 3HBS1

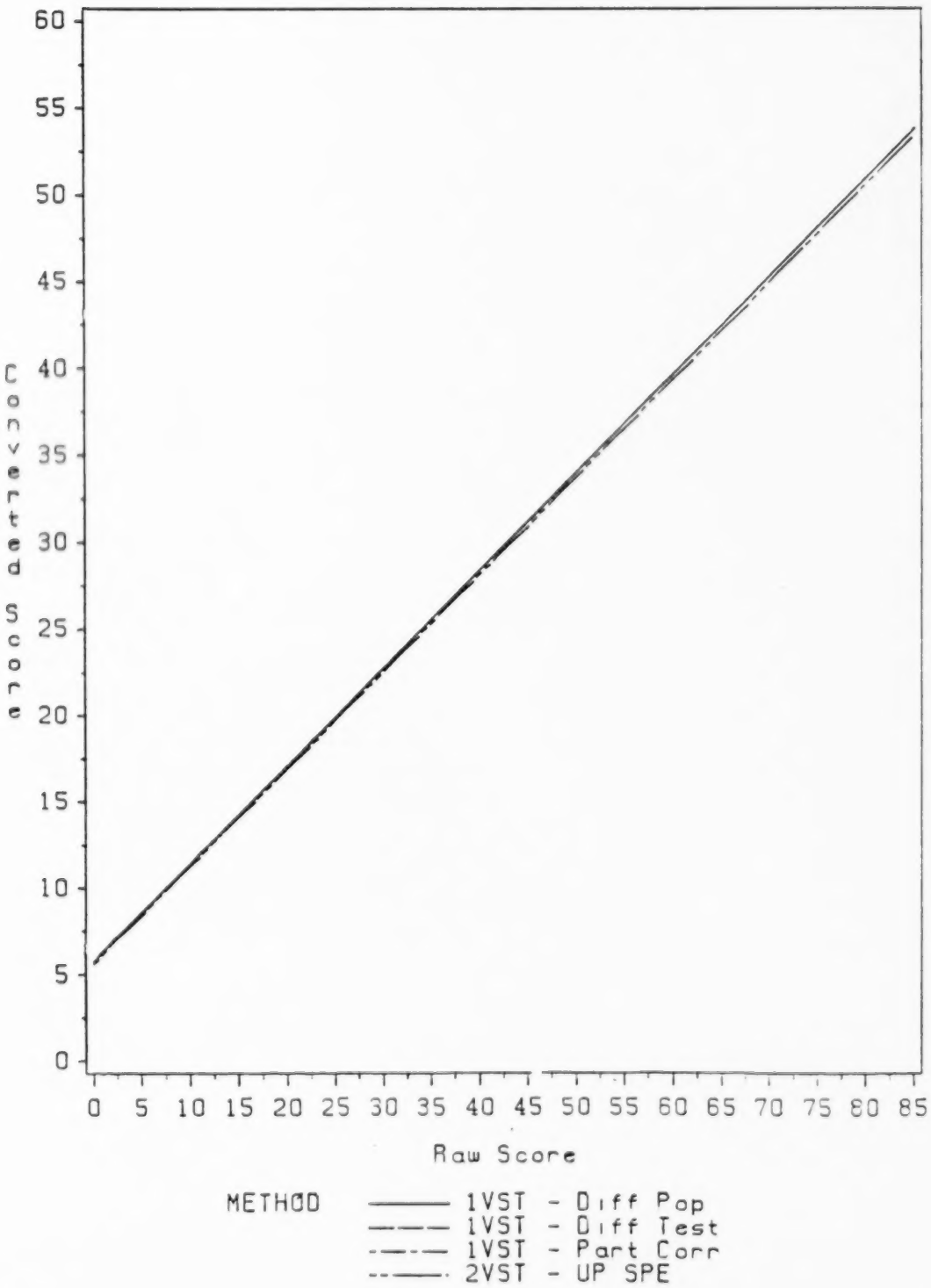


Figure 8

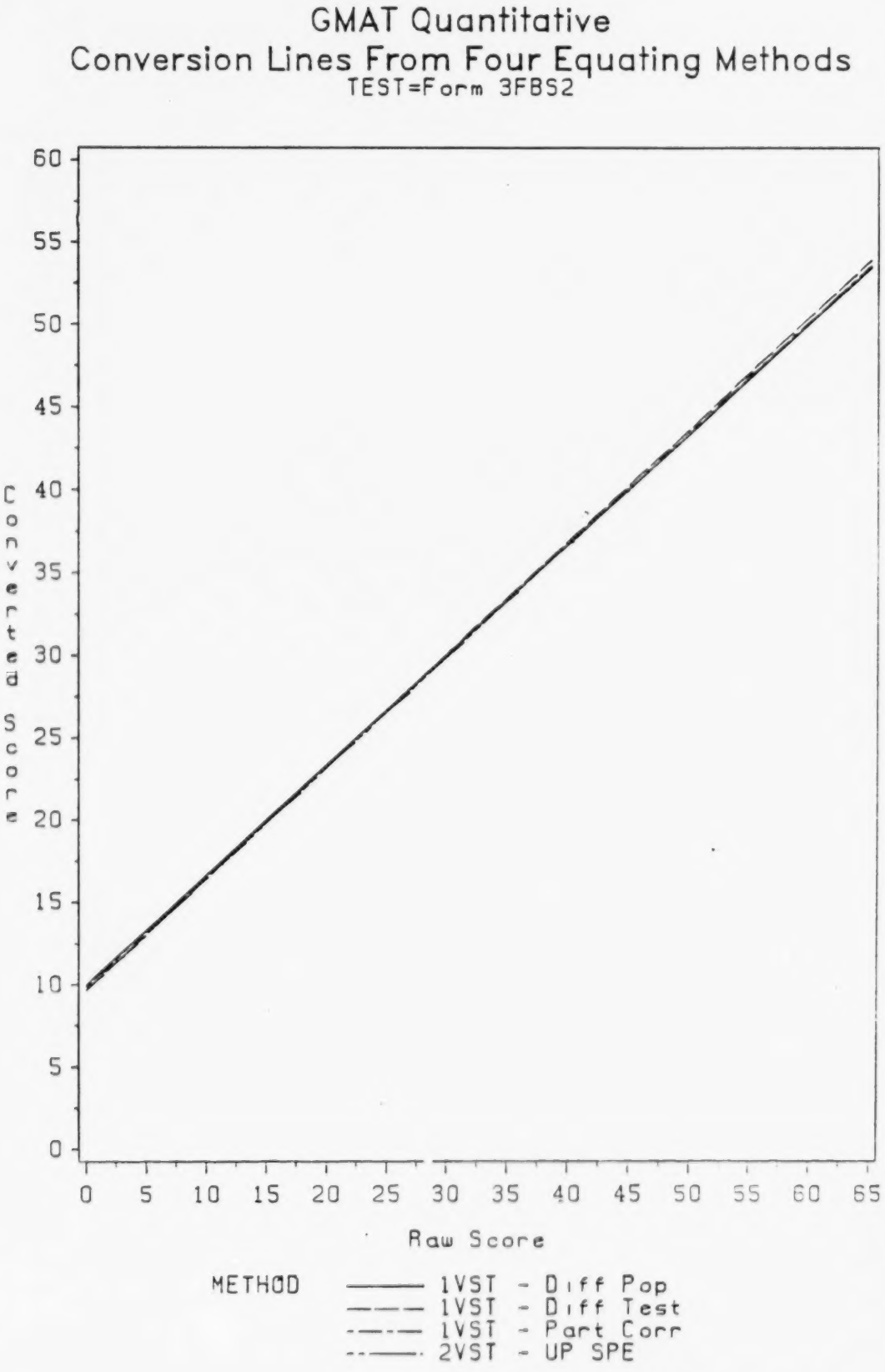

GMAT Quantitative
Conversion Lines From Four Equating Methods
TEST=Form 3FBS2
Converted Score
60
55
50
45
40
35
30
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Raw Score
METHOD
1VST - Diff Pop
1VST - Diff Test
1VST - Part Corr
2VST - UP SPE

Figure 9

GMAT Quantitative
Conversion Lines From Four Equating Methods
TEST=Form 3FBS4

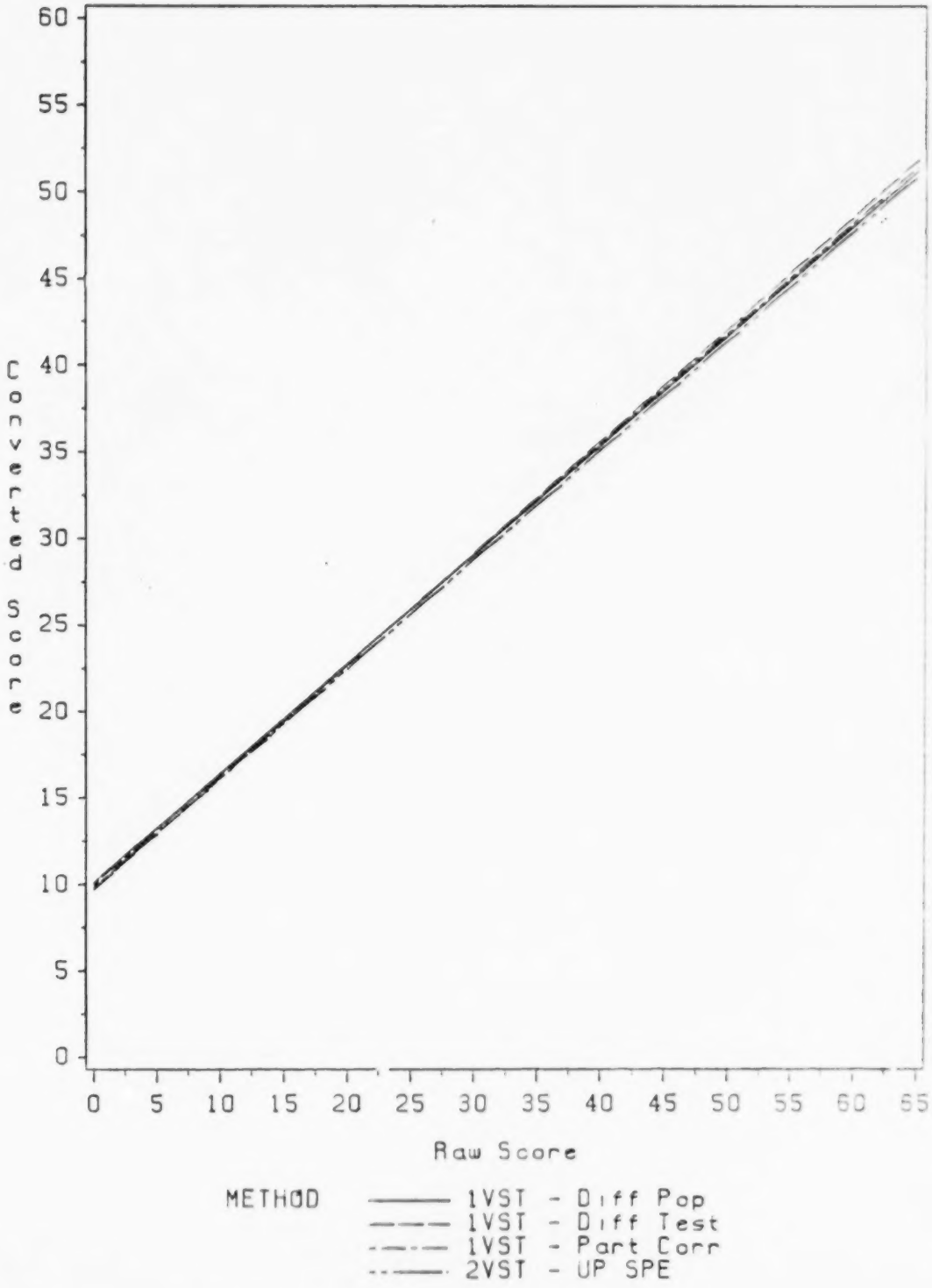


Figure 10

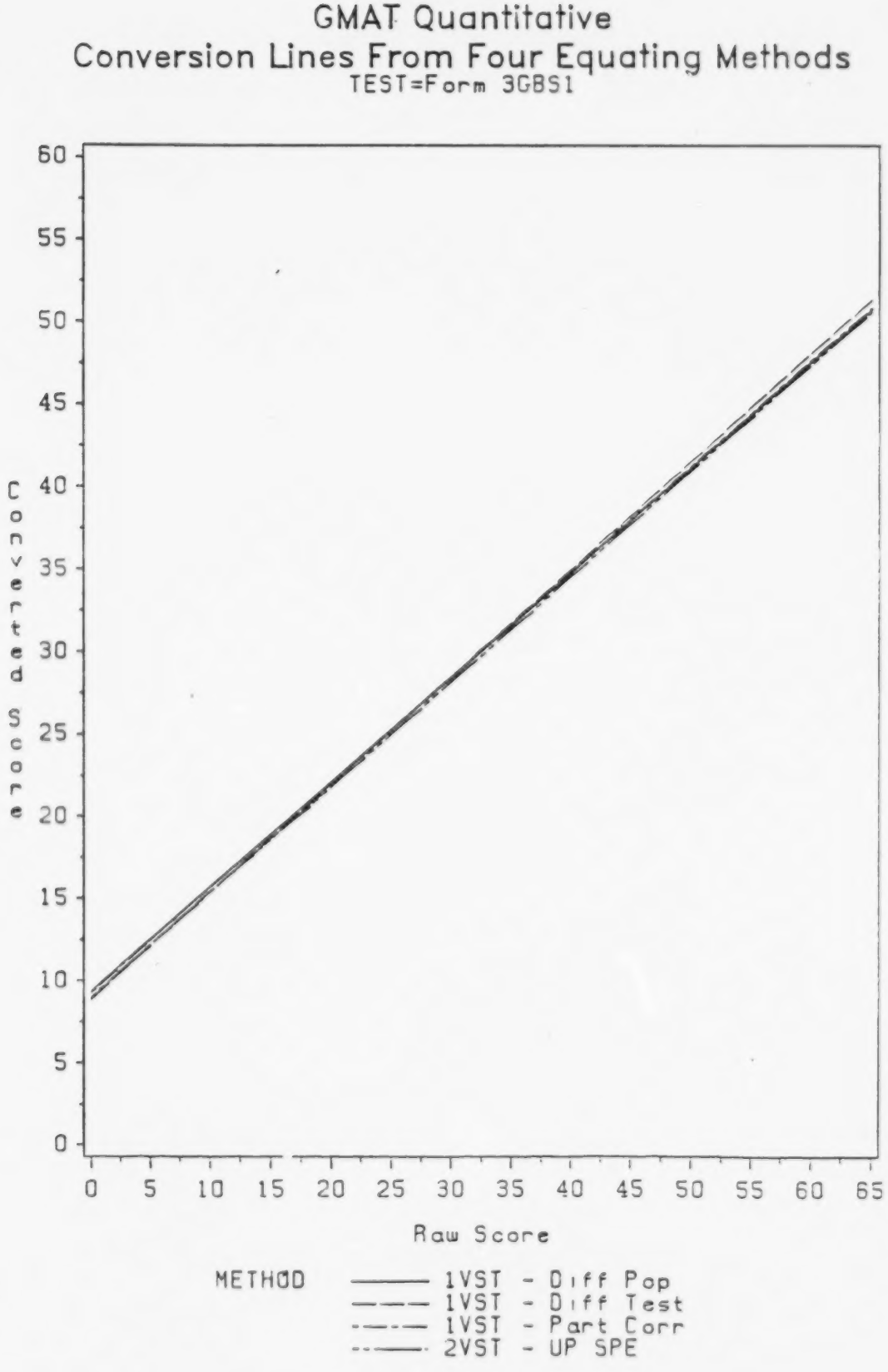

GMAT Quantitative
Conversion Lines From Four Equating Methods
TEST=Form 3GBS1
Converted Scale Score
60
55
50
45
40
35
30
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Raw Score
METHOD
1VST - Diff Pop
1VST - Diff Test
1VST - Part Corr
2VST - UP SPE

Figure 11

GMAT Quantitative
Conversion Lines From Four Equating Methods
TEST=Form 3GBS4

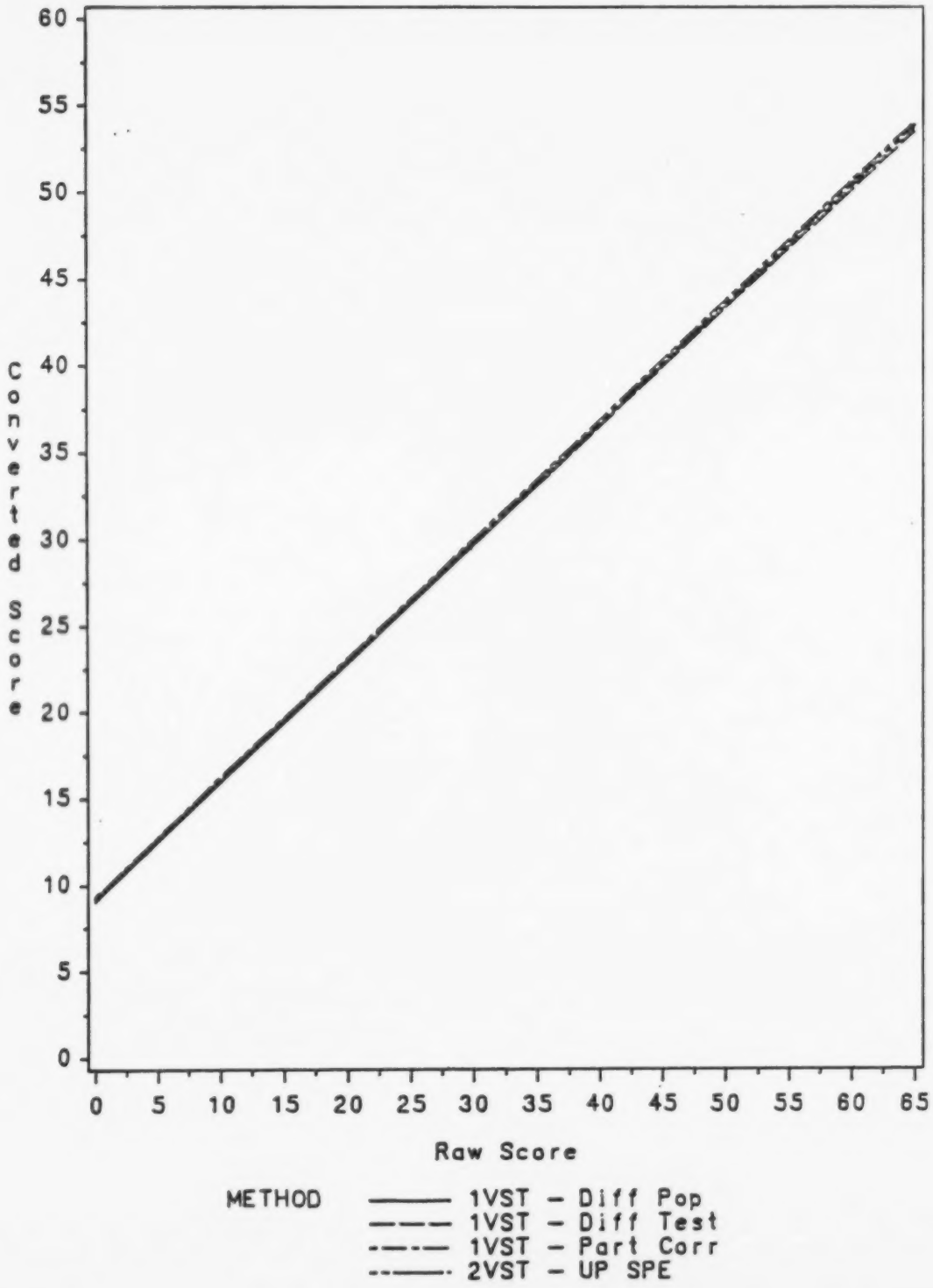
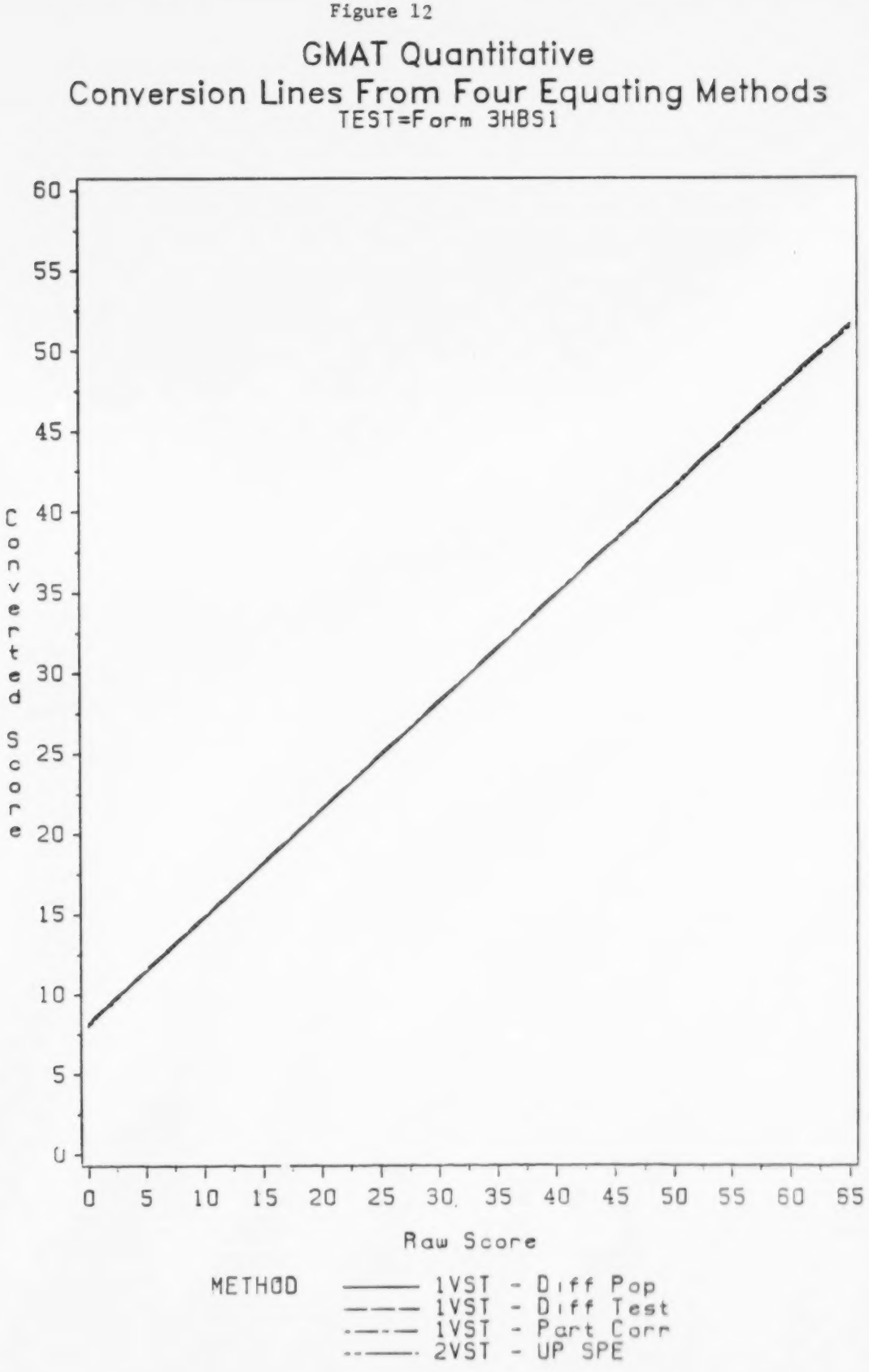


Figure 12


GMAT Quantitative
Conversion Lines From Four Equating Methods
TEST=Form 3HBS1
Converted Score
60
55
50
45
40
35
30
25
20
15
10
5
0
0 5 10 15 20 25 30 35 40 45 50 55 60 65
Raw Score
METHOD
1VST - Diff Pop
1VST - Diff Test
1VST - Part Corr
2VST - UP SPE

APPENDIX A

Preoperational section correlations independent of operational sections.

Verbal Partial Correlations

FORM	ADMN	.....R
AS/RC		
I3	JA85	0.1013
I3	M85	0.1195
I2	O84	0.1266
H2	O84	0.1368
I4	JA85	0.1424
I2	JA85	0.1467
H2	M84	0.1570
I4	JU85	0.1709
I1	JU84	0.2132
I1	O84	0.2373
AS/SC		
H2	M84	0.0979
H2	O84	0.0993
I4	JA85	0.1034
I2	O84	0.1138
I3	JA85	0.1188
I4	JU85	0.1285
I3	M85	0.1496
I2	JA85	0.1635
I1	JU84	0.1656
I1	O84	0.1750
RC/SC		
H2	O84	0.1204
I3	JA85	0.1433
H2	M84	0.1456
I4	JU85	0.1557
I3	M85	0.1754
I4	JA85	0.1828
I1	O84	0.1852
I2	JA85	0.1863
I2	O84	0.1962
I1	JU84	0.2189

Quantitative Partial Correlations

FORM	ADMN	.....R
PS1/PS2		
I2	M85	0.1440
I3	M85	0.1947
I4	JA85	0.2593
H2	M84	0.2616
I2	O84	0.2744
H2	O84	0.2943
I3	JA85	0.3036
I1	O84	0.3113
I1	JU84	0.3246
I4	JU85	0.3389
PS1/DS		
H2	O84	0.1285
I3	JA85	0.1337
I2	M85	0.1520
I3	M85	0.1535
H2	M84	0.1816
I2	O84	0.2180
I4	JU85	0.2260
I1	JU84	0.2344
I4	JA85	0.2397
I1	O84	0.2577
PS2/DS		
H2	O84	0.1335
I2	M85	0.1466
I3	JA85	0.1770
I3	M85	0.1780
I2	O84	0.1788
I4	JU85	0.1887
H2	M84	0.1953
I4	JA85	0.2051
I1	O84	0.2201
I1	JU84	0.2861